

9

Kaija-Liisa Nissinen

Tiedon hakijat
online - tietokantojen
peruskäyttäjinä

1996



ELÄKETURVAKESKUS
PENSIONSSKYDDSCENTRALEN

Monisteita
Papers

SISÄLLYSLUETTELO

| | |
|---|----|
| 1. Tiedon hakijat peruskäyttäjinä | 1 |
| 1.1. Yleistä | 1 |
| 1.2. Peruskäyttäjät tiedonhakutilanteessa | 2 |
| 1.3. Peruskäyttäjät kokeneina tiedon hakijoina | 5 |
| 1.4. Peruskäyttäjät ja hakutermien valinta | 6 |
| 1.5. Peruskäyttäjät ja Boolean-logiikka | 10 |
| 2. Miten peruskäyttäjät selviytyvät tiedonhaun pulmatilanteista? | 11 |
| 2.1. Batesin ideointitekniikat | 11 |
| 2.2. Batesin hakuteknikoita | 14 |
| 3. Lopuksi | 17 |

KIRJALLISUUTTA

LIITTEET

Eläketurvakeskus

25.4.1996

Kaija-Liisa Nissinen

1. Tiedon hakijat peruskäyttäjinä

1.1. Yleistä

Tietoa hakevat henkilöt tarvitsevat käytännössä joko laajaa kaiken kattavaa relevanttia tai vain muutamaa aiheeseen sopivaa (relevanttia) tietoa (Fidel 1991a, 490). Koska tiedonhakijoilla on kuitenkin rajoitettu kyky löytää uutta tietoa, niin he tarkoituksella seuraavat niitä tiedon jyväsiä, joita heillä jo ennestään on olemassa. Itse asiassa tiedonhaku liittyy ihmisen koko persoonaan, kokemukseen, tunteeseen ja toimintaan. (Kuhlman 1991, 362.)

Tietokantojen tuottajat ovat jo 1980-luvulta lähtien olleet kiinnostuneita siitä, miten tiedon hakijat etsivät tietoa. He ovat tutkineet mm., mitä peruskäyttäjät ovat löytäneet ja miten. (Fidel 1991a, 491.) Tässä kirjallisuuskatsauksessa pyritään selvittämään, mitä tutkimukset kertovat peruskäyttäjien hakutekniikasta sekä hakutermien käytöstä. Peruskäyttäjille tarjotaan katsauksen lopuksi myös tapoja, joita käyttämällä he voivat selviytyä tiedonhaun aikana syntyneistä ongelmatilanteista. Peruskäyttäjillä tarkoitetaan tässä kaikkia online-tietokantojen loppukäyttäjiä.

1.2. Peruskäyttäjät tiedonhakutilanteessa

1970-luvulla pääasiassa vain tiedonvälittäjät käyttivät hyväkseen tiedonhauissaan online-tietokantoja. Tietokantojen itsenäiskäyttö lisääntyi 1980-luvulla, kun tietokantojen tuottajat pystyivät tarjoamaan yhä helppokäyttöisempiä käyttöliittymiä tietokantojen käyttäjille. Tässä yhteydessä ryhdyttiin myös tutkimaan peruskäyttäjiä ja heidän tiedonhakuihin liittyviä valintojaan. Haluttiin esimerkiksi tietää, mitä kokemus tiedonhaussa vaikuttaa tiedonhakutekniikkaan ja hakutermien valintoihin.

Peruskäyttäjien tiedonhakutekniikka on tutkinut mm. Hsieh-Yee, jonka tutkimus rajoittui Dialog-järjestelmään ja sen ERIC-tietokantaan. Hän teki tutkimuksensa laboratorio-

oloissa niin, että kustannustekijät oli eliminoitu kaikista tiedonhakutilanteista. Ysieh-Yee valitsi koehenkilöiksi sekä vasta-alkajia että kokeneita tiedonhakijoita. **Kokeneeksi tiedon hakijaksi** tutkimuksessa on määritelty henkilö, joka on käynyt vähintään yhden online-tietokannan käyttökurssin ja jolla on vuoden ajalta hakukokemusta tiedonhausta. **Vasta-alkajana eli noviisina** on pidetty henkilöä, joka ei ole saanut koulutusta tiedonhaussa, ja joka on suorittanut alle viisi tiedonhakua. (Hsieh-Yee 1993, 162-164.)

Vasta-alkaja ei tiedonhakutilanteessa ole voinut käyttää hyväkseen haettavan aiheen tuntemuksen lisäksi syvempiä tietoja esimerkiksi hakutekniikasta kuin korkeintaan tiedon tuottajan tarjoamat alkeistiedot. Kokeneella tiedonhakijalla on sitävastoin haun aiheeseen liittyvien tietojen lisäksi tietoa itse tietokannasta ja siinä käytetystä kielestä. Koetilanteessa Hsieh-Yeella oli yhteensä 32 kokenutta tiedonhakijaa ja 30 noviisia. Tutkimustulokset on koottu tiedoista, joita kerättiin mm. koehenkilöitä havainnoimalla ja tehdyistä hakupöytäkirjoista (Hsieh-Yee 1993, 165). **Hakutekniikalla** (search tactics) tarkoitetaan tutkimuksessa haun aikana tehtäviä muutoksia, joiden avulla on tarkoitus päästä tiedonhaussa eteenpäin (Hsieh-Yee 1993, 161).

Kun Hsieh-Yee valitsi koetilanteeseen hakuaiheeksi sellaisen aiheen, jota tiedonhakijat eivät tunteneet, niin kokeneet peruskäyttäjät osasivat vasta-alkajia paremmin käyttää tiedonhaussa hyväkseen (Hsieh-Yee 1993, 161, 168-169.)

- omia hakusanojaan
- aiheeseen liittyvää tesaaurusta
- hakutermejä tietokannan ulkopuolelta
- hakutermien synonyymejä
- hakutermien yhdistämistä

Aiheen tuntemus on vaikuttanut positiivisesti vasta, kun peruskäyttäjällä on tietty määrä kokemusta tiedonhausta sinänsä (Hsieh-Yee 1993, 161-169).

Vasta-alkajat Siegfriedin yms. tekemän tutkimuksen mukaan pystyvät ajan kuluessa kehittämään tiedonhaun tekniikan hyväksikäyttäjinä. Tutkimuksen mukaan vasta-alkajat oppivat hyödyntämään hauissaan läheisyysoperaattoria, sanojen katkaisua ja sanojen ennalta spesifiointia yms. (Siegfried et al., 1993, 288.), mutta sulkujen käyttöä hakulausekkeissa tuskin esiintyi (Siegfried et al., 1993, 280).

Myös Sullivan yms. (1990, 28-30) on todennut tietokantojen peruskäyttäjistä, että tavalliset peruskäyttäjät ovat tyydyttäviä tiedonhakijoina, mutta tekevät enemmän virheitä ja vähemmän ennakkovalmisteluja kuin asiantuntijat (kokeneet tiedonhakijat). Saman tutkimusryhmän mukaan väitetään, että naiset tekevät miehiä useammin virheitä tiedonhaun aikana (Sullivan et al 1990, 27). Bellardo (1985a) on puolestaan todennut, että useimmat loppukäyttäjät tekevät virheitä.

Bellardo on keränyt peruskäyttäjistä tietoa havainnoimalla kuudessa eri informaatio-alan koulussa yhteensä 61:tä opiskelijaa. Hänen mielestään haun hyvää tasoa ei juurikaan voida ennustaa tiettyjen peruskäyttäjän persoonallisten ominaisuuksien perusteella. (Bellardo 1985b, 241.) Kokoamiensa tulosten perusteella hän on kaikkien peruskäyttäjien hakuteknikoista (hakutyyleistä) vetänyt edellisen lisäksi seuraavia johtopäätöksiä. (Bellardo 1985a, 236.):

- koulutuksen määrällä ja laadulla on huomattava vaikutus aloittelijan (vasta-alkajan) hakuihin, mutta saavutetut edut tasaantuvat aikaa myöten
- järjestelmään liittyvillä ominaisuuksilla paitsi kustannuksilla, ei näytä olevan juurikaan vaikutusta peruskäyttäjän käyttäytymiseen ja asenteisiin
- kokemuksella on vain vähän positiivista vaikutusta hakutuloksiin, mutta useat tutkijat ovat sitä mieltä, että saman kokemuksen omaavilla peruskäyttäjillä saattavat hakuteknikka ja hakutulokset vaihdella huomattavastikin

- niiden ominaisuuksien (matemaattisten, verbaalisten tms.), joiden on havaittu vaikuttavan positiivisesti yksilön ohjelmointitaitoihin, on havaittu tietyssä määrin (mutta ei välttämättä) vaikuttavan myös positiivisesti peruskäyttäjän hakutaitoihin
- ne peruskäyttäjät, jotka ovat joustavia ja luovia hakuongelmia ratkoessaan, pärjäävät hyvin myös itse tiedonhakutilanteessa. Näiden ominaisuuksien keskeinen riippuvuus ei kuitenkaan ole voimakas
- kaikki peruskäyttäjät toimivat interaktiivisimmin silloin, kun tiedonhaku tapahtuu tutussa tiedonhakuympäristössä ja normaalissa tiedonhakutilanteessa
- ne peruskäyttäjät, jotka ovat nopeita hakiessaan tietoa erissä (fast batch searchers), eivät ole huonompia tiedon hakijoina kuin interaktiivisetkaan loppukäyttäjät. Tietyissä tilanteissa he saattavat olla jopa tehokkaampia (esim. he saattavat löytää istunnon aikana enemmän relevantteja viitteitä minuutissa kuin muut tiedonhakijat)
- kokeneilla tiedon välittäjillä on selvästi muista peruskäyttäjistä erottuva hakutekniikka, joka saattaa vaikuttaa haun tulokseen positiivisesti, tai päinvastoin

Edellä **interaktiivisella tiedon hakijalla** tarkoitetaan peruskäyttäjää, joka pyrkii vaikuttamaan tiedonhaun tulokseen tekemällä runsaasti korjauksia yhden tiedonhaun aikana. Korrelaatiotestit osoittavat (Fidel 1991a, 520) interaktiivisen hakutyylin korreloivan subjektiivisten tekijöiden kanssa.

Tästä johtuen toiset peruskäyttäjät ovat interaktiivisempia kuin toiset. Sullivan yms. (1990, 27) väittävät vielä, että naiset ovat interaktiivisempia kuin miehet peruskäyttäjinä.

1.3. Peruskäyttäjät kokeneina tiedonhakijoina

Fidel (1984) on jakanut online-tietokantojen ammattikäyttäjät (asiantuntijat, kokeneet peruskäyttäjät) kahteen toisistaan selvästi erottuvaan ryhmään:

(a) operationalisteihin ja **(b)** konseptualisteihin.

Näistä ensimmäinen ryhmä **(a)** tarkoittaa niitä tiedon hakijoita, jotka hakuprosessin kestäessä tekevät hakulausekkeeseen muutoksia siten, että haussa mukana oleviin termeihin (vapaatermeihin tai asiasanoihin) ei vaikuteta asiasisältöä muuttamalla. Operationalistit käyttävät tiedonhaun muutostilanteissa hyväkseen pelkkää hakutekniikkaa. (Fidel 1985, 62 & Fidel 1984, 212-213.)

Operationalisti pyrkii supistamaan hakutulosta esim. julkaisuvuosien perusteella tai käyttämällä pelkkiä synonyymisanoja. Hän voi muuttaa viitejoukkoa myös etsimällä hakutermeille erilaisia kirjoitustapoja tai suorittamalla hakutermien katkaisuja. Tavallisesti hän valitsee hakutermeiksi vain sellaisia asiasanoja, jotka ovat jo aikaisempien tiedonhakujen avulla osoittautuneet hyvin aihetta kuvaaviksi hakutermeiksi. Näillä termeillä on löydetty aikaisemminkin ko. aihetta runsaasti. (Fidel 1984, 212-213.) Operationalisti uskoo tarjoavansa tietoa löytämällä sopivia lähteitä, joiden avulla päästään kiinni itse aiheeseen. Loppu jää tiedon käyttäjän omaksi huoleksi. (Fidel 1984, 218.)

Konseptualisti (b) sitävastoin pyrkii aluksi ymmärtämään hakukysymyksen yhden fasetin puitteissa valitsemalla kaikki asiasanat juuri tästä fasetista. Tiedonhaun aikana hän tekee muutoksia löydettyjen viitteiden sisältöön. Hän saattaa muuttaa kesken haun kyselylauseketta ja käytettyä indeksikieltä. (Fidel 1985, 62.) Hän saattaa muuttaa tiedonhakuun valitun termin laajempaan tai supistavaan termiin tai vaihtaa sen rinnakkaisen termiin. Hän voi lisäksi yhdistää toisesta fasetista jonkin elementin käytössä olevaan hakulausekkeeseen. (Fidel 1984, 213.) Konseptualisti uskoo löytäneensä melkein kaiken, mitä aiheesta tästä tietokannasta voikin löytää (Fidel 1984, 218).

Tämä asiantuntijoiden eli kokeneiden tiedon hakijoiden jakautuminen kahteen eri tiedonhakutyyppiin perustuu tutkimukseen, jossa havainnointiin viittä kokenutta tiedonhakijaa. Näistä kukin suoritti tutkimuksen kestäessä 10-13 tiedonhakua. Tiedonhaut analysoitiin vaihe vaiheelta aina tiedonhakuun valmistautumisesta, hakukysymyksen ymmärtämisestä, tietokannan ja termien valintaan saakka. Myös itse tiedonhakutilanteita ja lopputulosten kuvailua havainnointiin. (Fidel 1984, 213.)

1.4. Peruskäyttäjät ja hakutermien valinta

Tyypillinen online-tiedonhaku jakautuu suoritusten perusteella kolmeen edellä mainittuun vaiheeseen (Fidel 1991, 491). Fidel on 1990-luvulla keskittynyt tutkimuksissaan erikoisesti peruskäyttäjien hakutermien valintoihin. Hänen mielestään on olemassa kaksi hakutermityyppiä: vapaan tekstin termit (**vapaatermit**) luokittelematonta tekstiä varten, ja tesaauruksen **asiasanat** kontrolloitua tekstiä varten. Useissa tietokannoissa on mahdollista tehdä tiedonhaku hakemalla tietoa suoraan jollakin vapaatermillä tai vapaatermejä yhdistämällä tai asiasanoilla sekä asiasanoja yhdistämällä. Näiden termien käyttäminen samanaikaisesti haussa on myös mahdollista. (Fidel 1991a, 491.)

Asiasana määritellään Sf-standardin mukaan indeksointikielentermiksi, joka on valittu kuvaamaan tiettyä käsitettä. **Indeksointikieli** tarkoittaa dokumentaatiokieltä, jonka termit ovat luonnollisesta kielestä valittuja sanoja. (SFS 5471.1988,2.) Sf-standardin mukaan (SFS 5471.1988,7) **hakusanat** ovat tiedon haussa käytettäviä sanoja. Tässä katsauksessa hakusanasta käytetään myös nimitystä hakutermi tai termi. Hakutermi = asiasanat + vapaatermit (Fidel 1991c, 517).

Fidel on selvittänyt peruskäyttäjien hakutermien valintaa valitsemalla koehenkilöiksi 47 kokenutta peruskäyttäjää tai tiedonhaun ammattilaista. Tulokset on kerätty havainto-

ja kyselytutkimuksella sekä koehenkilöiden tekemillä hakupöytäkirjoilla. (Fidel 1991a, 490-492.) Termien valinnat tapahtuvat liitteessä 1 esitetyn "hakurutiinien" luettelon mukaisesti (Fidel 1991b, 503).

Lisäksi hakurutiinien luettelosta (Liite 1) selviää esimerkiksi, millä perusteella ja mistä käsitteeltään yleinen tai tarkasti rajattu (spesifi) hakutermi on valittu. Siitä näkyy myös, jos hakutermien rinnalle on valittu muita termejä tai samoja hakutermejä on käytetty toisessa tietokannassa. (Fidel 1991a, 493-495.)

Hakutermien frekvenssijakautumat on laskettu sen jälkeen, kun 47 koehenkilöä oli suorittanut yhteensä 281 hakua, joissa käytettiin yhteensä 3 200 termiä (Fidel 1991b, 502). Laskettujen jakautumien perusteella Fidel on kehittänyt muutamia tunnuslukuja (Fidel 1991b, 502, 504-505 & Fidel 1991b, 517). Ne ovat:

(1) Vapaatermien käyttökerroin (ilmaistaan %:na)

= kaikkien vapaatermien lukumäärä / kaikkien hakutermien lukumäärällä / = Lue:jaettuna

(2) Tesauruksen käytön hylkäyskerroin (ilmaistaan %:na)

= kaikkien niiden vapaatermien lukumäärä, joita perustäyttäjät eivät ole tarkistaneet asiasanastoista / kaikkien hauissa käytettyjen hakutermien lukumäärällä

(3) Vapaatermien ja asiasanojen lukumäärät

(4) Hauissa käytettyjen tietokantojen lukumäärä

(5) Istunnon aikana hakulausekkeeseen tehtyjen muutosten lukumäärä

(6) Eri aiheista tehtyjen tiedonhakujen lukumäärä

= peruskäyttäjää kohti tehtyt eri aihehakujen lukumäärät, eli osoittaa sen, mille aloille hakija on erikoistunut.

(Tässä voi esiintyä aiheita lääketieteen, luonnontieteen, sosiaalipolitiikan, humanististen ja yleistietojen alueelta)

(7) Tietoja peruskäyttäjän hakuympäristöstä käytännön, teoreettisella ja yleisellä tasolla

= Käytännön tasolla hakuympäristönä on laitos, jolle peruskäyttäjä suorittaa tiedonhakunsa. Teoreettisella tasolla hakuympäristönä on paikka, jossa suoritetaan tutkimuksia, kuten yliopistot, jokin toimisto jne. Yleisellä tasolla on kyse kaikista muista ympäristöistä.

Edellisten tunnuslukujen perusteella Fidel on voinut todeta, että kaikista koehenkilöiden käyttämistä termistä **puolet oli asiasanoja ja puolet hakusanoja**. Luvut kertovat edelleen, että kontrolloituja asiasanastoja eli tesauksia sekä aineiston indeksointia tarvitaan. Esimerkiksi tulokset osoittavat, että peruskäyttäjät tutkivat asiasanastoja kolme neljännestä termien valintaan kulutetusta ajasta. He myös kuluttivat puolet hakuun käytetystä ajasta hakemalla juuri asiasanoilla. (Fidel 1991b, 505.)

Ne peruskäyttäjät, jotka välttävät tiedonhauissaan asiasanojen käyttöä ilmoittivat vapaatermien suosimisen syyksi seuraavia asioita:

- he eivät luota asiasanastoihin tai indeksointiin
- heidän termsä eivät löydy asiasanastoista
- he suorittavat tiedonhaun useista eri tietokannoista
- heillä ei ollut asiasanastoa käytettävissään

Fidelin mukaan vastaukset osoittavat, että ainakaan huonosti laadittua asiasanastoa ei kannata käyttää. Tietenkään asiasanastoa ei voi käyttää, jollei sitä ole olemassa. (Fidel 1991c, 511.)

Lisäksi vastausten perusteella, kun tiedonhaku joudutaan tekemään useasta eri tietokannasta, niin tiedonhaussa vältetään asiasanaston käyttöä. Jos monien tietokantojen käyttäjillä olisi saatavilla tesaaurus, jossa asiasanalle annetaan "käännöstermi" sen mukaan, missä tietokannassa tätä käännöstermiä voidaan käyttää asiasanana, niin todennäköisesti tällaista moneen tietokantaa soveltuvaa tesaurusta tai termitietopankkia myös käytettäisiin. (Fidel 1991c, 513.)

Fidel on Pearsonin Moment-Product -korrelaatiotestin avulla osoittanut myös, että kokeneista tiedonhakijoista operationalistit käyttävät mieluummin vapaatermejä ja konseptualisti asiasanoja tiedonhauissaan. Vapaatermien käyttökertoimesta suoritettujen korrelaatiotestien perusteella Fidel päättelee lopuksi, että vapaatermien käyttö tiedonhaussa on riippuvainen seuraavista muuttujista (Fidel 1991c, 521-522.):

- tehtyjen operationaalisten muutosten määrästä
- tesauruksen käytön hylkäyskertoimesta
- aihealasta, eli onko kysymys yleisestä vai teoreettisesta aiheesta
- tieteenalasta, josta tietoa etsitään

Edellä mainitusta "tieteenalasta" Fidel mainitsee esimerkkinä lääketieteen ja luonnontieteen. Näissä tieteissä käytetyt sanat ovat jo sinänsä kontrolloituja termejä. Fidelin laskujen perusteella lääketieteessä peruskäyttäjät käyttävät tiedonhauissaan hyväksyen alan asiasanastoja useammin kuin luonnontieteessä. Luonnontieteessä peruskäyttäjät etsivät alan tietoja useimmiten vapaatermeillä. Lääketieteen piirissä tesaurukset ja indeksointi ovatkin tunnetusti hyvätasoisia, minkä perusteella Fidel olettaa, että tesauruksen käyttö lääketieteessä johtuu juuri lääketieteen piirissä huolella laadituista tesauruksista. (Fidel 1991c, 522.)

Siegfried & Bates & Wilde (1993) ovat havainneet, että humanististen alojen peruskäyttäjät eroavat huomattavasti muiden tieteenalojen peruskäyttäjistä. Tutkimuksen mukaan humanistit hakevat muiden tieteenalojen edustajia useammin tietoa henkilöiden nimen, maantieteellisten termien ja ajanjakson mukaan. Lisäksi he käyttävät run-

saasti oman oppialansa termejä. Tiedot on kerätty Getty Centerissä vuosina 1988-1990 humanisten alojen tiedemiesten tiedon hakujen tuloksista. He ovat käyttäneet Dialog -järjestelmää tietojensa hankkimiseen. (Bates & Wilde & Siegfried 1993, 1, 3-4, 30.)

1.5. Peruskäyttäjät ja Boolean-logiikan käyttö

Kun hakukysymyksestä muodostetaan ensin hakulauseke, niin se muodostetaan eri hakuelementeistä (hakutermeistä) ja Boolean operaattoreista. Nämä hakuelementit ovat joko yksittäisiä sanoja tai sanaliittoja (Bates et al 1993, 27.). Boolean operaattoreina käytetään AND, OR tai NOT -operaattoreita. Näistä AND leikkaa, OR ryhmittää ja NOT erottaa hakutermejä toisistaan. (Fidel 1985, 61.) Tällä Boolean-logiikan käytöllä tiedon haussa on havaittu olevan eroja vasta-alkajien ja kokeneiden käyttäjien kesken. Edellisessä Dialog-järjestelmää tutkivassa tiedonhakututkimuksessa (Getty Art History Information Program) kiinnitettiin huomiota vasta-alkajien Boolean operaattoreiden käyttöön hakulausekkeissa.

Vertailuryhmänä tässä tutkimuksessa oli asiantuntijaryhmä, joka työkseen teki tiedonhakuja Dialogi-järjestelmässä. Tämän asiantuntijaryhmän tekemiin hakuihin verrattuna vasta-alkajien Boolean operaattoreiden käyttö pysyi vielä yksinkertaisena. Vaikka koehenkilöiden tiedot hakutekniikasta ja Boolean operaattoreista kokeen kestäessä (3 vuotta) koko ajan karttuivat, niin kokeen loppuvaiheessakin heidän tiedonhauistaan peräti 63 %:a oli hakuja, jotka suoritettiin yhdellä hakutermillä. Ehkä he osasivat kuitenkin leikkaa ns. settejä AND-operaattorilla paremmin kuin aikaisemmin (Siegfried et al., 1993, 279-280). Sitäpaitsi neljännes koeryhmään kuuluvista vasta-alkajista ei käyttänyt kertaakaan OR-operaattoria tiedonhauissaan. (Siegfried 1993 et al., 1993, 288).

Edelliset tulokset osoittavat, kuinka Boolean logiikan käyttö on ja pysyy ongelmana peruskäyttäjille, jos he eivät ole tiedonhaun ammattilaisia. Samalla lailla ovat sanoneet mm. Misch Lee 1987 ja Sullivan et al. 1990 (Siegfried et al., 1993, 288).

2. Miten peruskäyttäjät selviytyvät tiedonhaun pulmatilanteista?

Tiedonhaussa **hakustrategialla** tarkoitetaan sitä suunnitelmaa, joka tehdään koko tiedonhakua varten, eli siinä suunnitellaan hakulausekkeen muoto ja mukaan otettavat hakutermiit. Muutoksilla itse hakutilanteessa hakija pyrkii vaikuttamaan jo tehtyyn hakuun (Fidel 1985, 61).

Bates (1979a) on koonnut tutkimuksen kautta tietoa edellä mainitusta, hakutilanteessa tehtävistä muutoksista. Hän on löytänyt 29 hakutekniikkaa, joilla on tarkoitus muuttaa itse hakutilannetta. Lisäksi hän on erottanut 17 toisistaan selvästi eroavaa ideointitekniikkaa, joilla voidaan ratkaista hakutilanteessa syntyviä ongelmia. Näiden tarkoituksena on vaikuttaa peruskäyttäjän ajatteluprosessiin tiedonhaun aikana. Seuraavassa esitetään luettelo näistä ideointiteknikoista (idea tactis). Lopuksi esitetään eräitä Batesin mainitsemia hakuteknikoita. Tarkoituksena on teknikoiden esittämisellä auttaa peruskäyttäjää toimimaan ongelmatilanteessa ideaalisella tavalla. (Bates 1979b, 280.) Nämä tekniikat soveltuvat sekä biograafiseen että online-tiedonhakuun.

2.1 Batesin ideointitekniikat

17 Batesin mainitsemaa ideointitekniikkaa ovat (Bates 1979b, 282-285.):

1. THINK (ajattelu)

- peruskäyttäjä pystyy oman ajattelun avulla löytämään uusia ideoita ongelmatilanteessa

2. BRAINSTORM (ideatulva, neronleimaus)

- peruskäyttäjän ajatuksissa syntyy nopeasti useita uusia ideoita, joita kohtaan hän ei saa olla liian kriittinen vaan hänen tulee käsitellä kaikki syntyneet ideat kunnes löytää niistä hyvin muotoillut ja toteuttamiskelpoiset tapaukset

3. MEDITATE (mietiskely)

- peruskäyttäjän rationaaliset ja intuitiiviset voimavarat saatetaan yhteistyöhön ratkaisemaan olemassaolevaa ongelmaa

4. CONSULT (neuvojen hakeminen)

- peruskäyttäjä pyytää ehdotuksia kollegoiltaan

5. RESCUE (pelastusoperaatio)

- peruskäyttäjä tutkii mahdollisesti käyttämättä jääneet reitit ja kokeillaan niitä tiedonhakuun

6. WANDER (vaellus)

- peruskäyttäjä siirtyy lähteeltä toiselle ja yhdistää jo löydettyjä viitteitä, ja käyttää hyväkseen sattumalta muualta löydettyjä hyödyllisiä vihjeitä

7. CATCH (sieppaus)

- peruskäyttäjä muuttaa tiedonhankintaansa esimerkiksi lainaamalla puuttuvat elementit muista lähteistä tai kirjastoista

8. BREAK (keskeytys)

- peruskäyttäjä keskeyttää tiedonhaun ainakin tilapäisesti ja tekee uuden aiheeseen paremmin sopivan hakusuunnitelman

9. BREACH (läpimurto)

- peruskäyttäjä lähtee ihan uusille alueille, joissa sisältö vastaa paremmin tiedonhaun aihetta, eli vastaus löytyy ihan ennen käyttämättömistä, älyllisistä tai fyysisistä lähteistä. Tässä peruskäyttäjä hylkää tutun hakuympäristön.

10. REFREME (uudelleen suunnittelu)

- peruskäyttäjä tekee uudet raamit tiedonhauille sen mukaan, mitä informaatiota hän aiheesta on saanut tiedonhaun aikana muiden kanssa aiheesta keskusteltuaan. Hän pyrkii välttämään tiettyyn aiheeseen liikaa keskittyvän (vinoutuneen) hakuaiheen.

11. NOTICE (neuvojen käyttö)

- peruskäyttäjä seuraa neuvoja ja tietoja, joita hän on saanut hakukysymyksen luonteesta tai tiedonvälittäjän antamasta informaatiosta

12. JOLT ("yllätysiskut")

- peruskäyttäjä muuttaa tiedonhakutilanteen täysin erilaiseksi lähestymällä ongelmaa aivan erilaisella, dramaattisella ajatustavalla

13. CHANGE (suunnitelmanmuutos)

- peruskäyttäjä muuttaa jotakin alkuperäiseen hakuun liittyvää termiä, tietolähdettä tms.

14. FOCUS (keskittää)

- peruskäyttäjä siirtyy käsittelemään pala palalta aihetta ja luopuu hallitsemasta kokonaisuutta kerralla

15. DILATE (laajentaa)

- peruskäyttäjä siirtyy hakemaan laajempaa kokonaisuutta, esim. siirtyy käyttämään suppeiden termien sijasta laajoja aiheeseen liittyviä termejä

16. SKIP (hyppelehtiä)

- peruskäyttäjä keskeyttää monimutkaisen kysymyksen käsittelyn ja hakee aiheeseen liittyviä asioita sieltä täältä ja yksinkertaisemmasta näkökulmasta

17. STOP (lopetus)

- peruskäyttäjä keskeyttää väliaikaisesti tiedonhaun. Hän siirtyy toistaiseksi muiden töiden pariin

2.2. Batesin hakuteknikoita

Edellä mainituista ideateknikoiden lisäksi voidaan haun aikana tehdä muitakin muutoksia. Esim. muutostilanteita syntyy, kun (Fidel 1985, 62.)

- löytynyt viitejoukko on liian pieni
- löytynyt viitejoukko on liian suuri
- löytynyt viitejoukko on epärelevanttia

Silloin, kun **on löytynyt liian pieni viitejoukko**, niin Bates (1979a, 208, 212-213) on ehdottanut seuraavia tapoja suurentaa tulosta:

(a) SUPER (laajentaa hakua)

- peruskäyttäjä siirtyy hakemaan termeillä, jotka ovat tesaauruksen hierarkiassa hakutermiin laajempia termejä

(b) RELATE (hakea rinnakkaistermeillä)

- peruskäyttäjä siirtyy hakemaan termeillä, jotka ovat tesaauruksen hierarkiassa hakutermiin nähden rinnakkaisia termejä

(c) REDUCE (madaltaa hakukynnystä)

- peruskäyttäjä yksinkertaistaa tai löyhentää hakua mm. pudottamalla hausta pois AND
- operaattorilla rinnastetut termit

(d) PARALLEL (lisää synonyymejä)

- peruskäyttäjä lisää hakuun hakutermien synonyymejä tai hakutermien käsitteeseen läheisesti rinnastettavissa olevia termejä

(e) NEIGHBOR (lisää aihetta sivuavia termejä)

- peruskäyttäjä lisää hakuun vapatekstin termejä tai aikaisemmin käytettyjä termejä, jotka läheisesti sivuavat aihetta

(f) TRACE (jäljittää termejä)

- peruskäyttäjä tutkii ja jäljittää termejä, joilla on aikaisemmin löytänyt ko. aihetta, ja lisää näitä hakuun

(g) VARY (muuntaa termejä)

peruskäyttäjä vaihtaa termejä muuntamalla alkuperäisiä hakutermejä

Silloin, kun **on löytynyt liian suuri viitejoukko**, niin

Bates (1979a, 208, 212) on ehdottanut seuraavia tapoja pienentää tulosta:

(aa) SUB (käyttää hakutermien sijasta termille annettuja supistavia termejä)

(bb) EXHAUST (käyttää AND-operaattoria ja tiukentaa hakua)

(cc) PINPOINT (vähentää synonyymi- tai rinnakkaistermien määrää)

(dd) BLOCK (sulkee pois tietyt termit NOT-operaattorilla, eli hylkää tietyn käsitteen)

Bates (1979a, 208-209) on ehdottanut myös seuraavaa valvontaohjelmaa, jolla vältetään jo ennalta mahdollinen epärelevantti tulos:

- verrataan alkuperäistä hakukysymystä hakulausekkeeseen ja sillä tulostettuun viitejoukkoon ja todetaan siinä esiintyvät relevantit viitteet (CHECK)
- tutkitaan haun rakenne ja suunnitellaan se uusiksi, jos alkuperäinen malli ei vaikuta tehokkaalta (PATTERN)
- korjataan kirjoitusvirheet ja hakuaiheen kanssa ristiriidassa olevat termivalinnat (CORRECT)
- huolehditaan käytettyjen polkujen muistiin merkitsemisestä, etteivät jäljet katoa, jotta myöhemmin voi esim. täydentää jotakin jo käytettyä hakureittiä tms. (RECORD)

Eroa ei ole voitu löytää niiden tehtyjen muutosten (hakuteknikoiden) kesken, joilla on tarkoitus vähentää tai lisätä viitejoukkoa ja samalla parantaa tarkkuutta tai saantia. Kun tiedonhakija tekee tietyn muutoksen tiedonhaussaan, niin hänen oma panoksensa vaikuttaa siihen, kumpaanko - saantiin tai tarkkuuteen - muutos kohdistetaan. (Fidel 1985, 63.) **Saanti** ilmaisee järjestelmän kykyä hakea relevanttia (sopivaa) aineistoa kokoelmasta. **Tarkkuus** ilmaisee järjestelmän kykyä torjua tuloksesta ei-toivottu aineisto. (Su, 1994, 207.)

Käytännössä peruskäyttäjällä on tarkoitus parantaa relevanttien viitteiden löytymistä. (Fidel 1991b, 505-506 & Su 1994, 7.) Hakutuloksen hyvät tarkkuusarvot eivät kuitenkaan aina takaa hyvää tiedonhaun laadullista tulosta. Esimerkiksi hakutulos, jolle on laskettu korkea tarkkuusarvo saattaa sisältää useita aikaisemmin tunnettuja, ei-toivottuja viitteitä haetusta aiheesta. Toisaalta taas hakutulos, jolle on laskettu edellistä matalampi tarkkuusarvo, saattaa sisältää juuri ne viitteet, joita tiedon hakija on etsinytkin. (Su 1994, 216-217.) Lopulta vain tiedon tarvitsija itse pystyy arvioimaan, miten tyydyttävänä hakutulosta voidaan pitää.

3. Lopuksi

Tästä katsauksesta selviää, että online-tietokantojen peruskäyttäjät jakautuvat vastaalkajiin ja kokeneisiin tiedon hakijoihin. Kokeneiden joukosta löytyy edelleen operationalisteja ja konseptualisteja. Katsauksessa pyritään kuvaamaan kaikkien peruskäyttäjien hakutyylejä eli -tekniikkaa sekä hakutermien valintoja. Hakutekniikasta mm. Bellardo on todennut, että koulutuksen määrällä ja laadulla on huomattava vaikutus vastaalkajien hakutekniikkaan, mutta saavutetut edut tasaantuvat aikaa myöten.

Fidel puolestaan on todennut, että online-tiedonhauissa peruskäyttäjän hakutermeistä puolet on asiasanoja ja puolet vapaatermejä. Lisäksi Fidelin tutkimukset osoittavat, että tesauroksia tarvitaan. Hänen mukaansa tesauroksia myös käytetään, jos ne ovat hyvin laadittuja. Tosin operationalistit käyttävät mieluummin vapaatermejä ja konseptualistit asiasanoja tiedonhauissaan.

Katsauksen lopussa esitetään lukijalle Batesin 17 ideointitekniikkaa ja eräitä tiedonhakuteknikoita. Peruskäyttäjä voi huoletta turvautua Batesin hakuteknikoihin, joiden tarkoituksena on auttaa tiedon hakijaa selviytymään pulmatilanteista. Kun vertaa Batesin hakuteknikoita ja Fidelin koehenkilöiden hakutermien valintarutiineja toisiinsa, lukija voi löytää yhtäläisyyksiä (Vrt. Liite 1).

KIRJALLISUUTTA

Bates, M.J. 1979a. Information Search Tactics. *Journal of the American Society Of Information Science* July 1979, 205-214.

Bates, M.J. 1979b. Idea Tactics. *Journal of the American Society for Information Science* September 1979, 280-289.

Bates, M. J. & Wilde, D. N. & Siegfried, S. (1993) An Analysis of Search Terminology used by Humanities Scholars: The Getty Online Searching Project. Report No. 1. *Library Quarterly* 63:1993, 1-39.

Bellardo, T. (1985). An Investigation of Online Searchers. Traits and Their Relationship to Search Outcome. *Journal of the American Society for Information Science* 36(4):1985, 241-250.

Bellardo, T (1985a) What Do We Really Know About Online Searchers? *Online Review* 9:1985, 223-239.

Bellardo, T. (1985b) An Investigation of Online Searchers. Traits and Their Relationship to Search Outcome. *Journal of the American Society for Information Science* 36(4):1985, 241-250.

Fidel, R. (1984). Online Searching Styles: A Case-Study-Based Model of Searching Behavior. *Journal of the American Society for Information Science* 35(4)1984, 211-221.

Fidel, R. (1985). Moves in Online Searching. *Online review* 9(1):1985, 61-74.

Fidel, R. (1991a). Searchers' Selection of Search Keys: The Selection Routine. *Journal of the American Society for Information Science* 42(7):1991, 490-500.

Fidel, R. (1991b). Searchers' Selection of Search Keys: II Controlled Vocabulary of Free-Text Searching. *Journal of the American Society for Information Science* 42(7):1991, 501-514.

Fidel, R. (1991c). Searchers' Selection of Search Keys: III Searching Styles. *Journal of the American Society for Information Science* 42(7):1991, 515-527.

Hsieh-Yee, I. (1993). Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science* 44(3):1993, 161-174.

Kuhlthau, C.C. 1991. Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science* 42(5):1991, 361-371.

SFS 5471. 1988. Suomenkielisen tesaauruksen laatimis- ja ylläpito-ohjeet. Tietopalveluseura 1988. 10 s.

Siegfried, S. & Bates, M.J. & Wilde, D.N. (1993). A Profile of End-User Searching Behavior by Humanities Scholars: The Getty Online Searching Project Report No. 2. *Journal of the American Society for Information Science* 44(5):1993, 273-291.

Su, L.T. (1994). The Relevance of Recall and Precision in User Evaluation. *Journal of American Society for Information Science* 45(3):1994, 207-217.

Sullivan, M. V. & Borgman, C. I. & Wippert, D. (1990) End-Users, Mediated Searchers, and Front-End Assistance Programs on Dialog: A Comparison of Learning, Performance, and Satisfaction. *Journal of the American Society for Information Science*, 41:1990, 27-42.

Liite 1/1. Hakutermien valintarutiinit



Käsite ei ole

uskottavasti

indeksitermi

1. Käyttää vapaasanoja

Käsitteellä on

useita synonyymejä

2. Käyttää asiasanoja

Loppukäyttäjä ei

ymmärrä

käsitteen merkitystä

3. Käyttää asiasanoja

Käsitettä ei ole

ehkä mainittu

tarkasti

4. Käyttää asiasanoja

Saantia on

parannettava

5. Liittää vapaasanoja
synonyymeinä asiasanoihin

6. Lisää seuravaaksi laajemman
hierarkiassa esiintyvän termin

7. Käyttää yleistä asiasanaa
joka sisältää koko käsitteen

Tarkkuutta on

parannettava

8. Rajaa haun vain asiasanoihin

9. Rajaa haun tärkeimpiin
asiasanoihin

10. Lisää rooli-indikaattoreita

11. Spesifioi dokumentin

12. Käyttää vapaasanojen
synonyymejä yhdessä
asiasanojen kanssa

Haku on

suoritettava

useasta tieto-

kannasta

13. Käyttää asiasanoja
vapaasanoina toisessa
tietokannassa



ELÄKETURVAKESKUS
PENSIONSSKYDDSCENTRALEN

PL 11, 00521 HELSINKI
Puh. 90-1511 Fax 90-1512496

PB 11, 00521 Helsingfors

P.O.Box, FIN-00521 Helsinki, Finland,
Tel. +358-0-1511, fax +358-0-151 2496

Eläketurvakeskus on Suomen työeläkejärjestelmän lakisääteinen keskuslaitos. Sen tutkimustoiminta koostuu pääasiassa sosiaaliturvaan ja eläkejärjestelmiin liittyvistä aiheista. Tutkimuksissa pyritään monipuolisesti ottamaan huomioon sosiaalipoliittiset, sosiologiset ja taloudelliset näkökulmat.

Pensionsskyddscentralen är lagstadgat centralorgan för arbetspensionssystemet i Finland. Forskningsverksamheten koncentrerar sig i huvudsak på den sociala tryggheten och på de olika pensionssystemen. Målet för forskningsprojekten är att mångsidigt belysa aspekter inom socialpolitik, sociologi och ekonomi.



THE CENTRAL
PENSION SECURITY INSTITUTE

The Central Pension Security Institute is the statutory central body of the Finnish employment pension scheme. Its research activities mainly cover the fields of social security and pension schemes. The studies aim to paint a comprehensive picture of the sociopolitical, sociological and financial aspects involved.

ISSN 1238 – 4984